LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO –
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type T060089-00-Z 03 May 2006
Technical Note

Point estimate variability in stochastic background searches due to windowing and notching effects

Albert Lazzarini

California Institute of Technology
LIGO Laboratory - MS 18-34
Pasadena CA 91125
Phone (626) 395-212
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Laboratory - MS NW17-161
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

www: http://www.ligo.caltech.edu/

Is it a Systematic or a Statistical Error?

Albert Lazzarini 2005.04 .08

■ Analysis of variability of the stochastic background point estimate vis-a-vis its corresponding estimated error

■ Background

The stochastic background point estimate has been observed to vary by a substantial (but much less than 1) fraction of its corresponding standard deviation when the same stretch of data are re-analyzed under slightly different conditions. It has been asked whether this is indicative that there is an underlying source of systematic error that is not being accounted for.

This simple analysis of a white noise (Gaussian) data stream provides insight into how much slightly different analysis conditions may affect answers that are derived from the statistics of white noise.

■ Statement of the problem

One source of varying point estimate that has been observed is the effect on the point estimate that the removal of, say, 1, 3, 5 or more frequency bins around known line harmonics has on the answer. Typically, the stochastic cross-correlation spectra are measured with 0.25 Hz resolution over a band of ~250 Hz $\Longrightarrow$ an analysis that involves integrating or summing together ~1000 noise-dominated frequency bins of spectrum to produce a point estimate. How much can this estimate vary if ~10% of the bins are dropped from an analysis?

■ Model problem

Assume we form an estimate ***y*** using N uncorrelated data points, x_i , each of which is a random variable taken from a Gaussian PDF:

$$\langle x_i \rangle = 0;\ \langle x_i\, x_j \rangle = \delta_{ij}\, \sigma_x^2$$

$$y(N) \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\langle y(N) \rangle = 0;\ \langle y(N)^2 \rangle = \left\langle \frac{1}{N^2} \sum_{i,j=1}^{N} x_i\, x_j \right\rangle = \frac{N\, \sigma_x^2}{N^2}$$

$$\sigma_{y(N)}^2 = \frac{\sigma_x^2}{N}$$

Now, what happens if p data points are dropped from the estimate Y(N), i.e., Y(N)$\longrightarrow$Y(N-p)?
We know from above that,

$$\langle y\,(N-p)\rangle = 0;\ \langle y\,(N-p)^2\rangle \ = \left\langle \frac{1}{(N-p)^2} \sum_{i,j=1}^{N-p} x_i\, x_j \right\rangle$$

$$\sigma^2_{y\,(N-p)} = \ \frac{\sigma_x^2}{N-p}$$

Now, however, consider the statistics of the difference $\delta y(p) \equiv y(N) - y(N-p)$: how much can the individual point estimates vary from one another?

$$\langle \delta y\,(p)\rangle \ = \ \langle y\,(N) \ - \ y\,(N-p)\rangle = 0;$$

$$\langle \delta y\,(p)^2\rangle \ = \ \langle y\,(N)^2\rangle \ - 2\,\langle y\,(N)\,y\,(N-p)\rangle \ + \langle y\,(N-p)^2\rangle$$

$$= \ \frac{\sigma_x^2}{N} \ + \ \frac{\sigma_x^2}{N-p} - 2\,\frac{1}{N\,(N-p)} \sum_{i=1}^{N} \sum_{j=1}^{N-p} \langle x_i\, x_j\rangle$$

$$= \ \frac{\sigma_x^2}{N} \ + \ \frac{\sigma_x^2}{N-p} - 2\,\frac{\sigma_x^2}{N}$$

$$= \ \sigma_x^2 \left(\frac{1}{N-p} \ - \ \frac{1}{N}\right) \approx \left(\frac{p}{N}\right) \frac{\sigma_x^2}{N},$$

$$\text{recalling}\ \sigma_y^2 \ = \ \frac{\sigma_x^2}{N},$$

$$\sigma_{\delta y} \ \approx \ \sqrt{\frac{p}{N}}\ \sigma_y$$

So, for example, assume we compare the shift in point estimate that can be observed when we compare an analysis performed with 1 vs. 3 bins dropped at each of the harmonics of 60Hz and 16 Hz in the spectrum from 50 Hz to 315 Hz. This frequency band contains 16 harmonics of 16 Hz are: {64, 80, 96, 112, 128, 144, 160, 176, 192, 208, 224, 240, 256, 272, 288, 304}. In addition it contains 5 harmonics of 60 Hz: {60, 120, 180, 240, 300}. The band {50, 315} Hz contains N = 1061 0.25 Hz bins. Notching each harmonic by a mask of width 1 bin removes p1 = 21 bins. The starting spectrum thus has N = 1040 bins; notching each harmonic with a mask of width 3 bins removes 63 bins, or an additional $\Delta p = p3 - p1 = 42$ bins; How much can we expect the point estimate y(1040) and y(998) to vary from each other based on the statistics of Gaussian noise?

$$\sigma_{\delta y} \ \approx \ \sqrt{\frac{\Delta p}{N}}\ \sigma_y \ = \ \sqrt{\frac{42}{1040}}\ \sigma_y \ = \ 0.2\,\sigma_y$$

Thus, it can be seen that only removing 4% of the frequency bins can cause the point estimate to shift by 20% of its corresponding error bar. Of course, the variances of y(1040) and y(998) are very similar:

$$\sigma_{y\,(1060)} \ \approx \ \sqrt{\frac{1}{1060}}\ \sigma_x \ = \ 0.0307\,\sigma_x$$

$$\sigma_{y\,(998)} \ \approx \ \sqrt{\frac{1}{998}}\ \sigma_x \ = \ 0.0313\,\sigma_x$$